

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problems Mailbox.**

3/9/1  
DIALOG(R) File 347:JAPIO  
(c) 2000 JPO & JAPIO. All rts. reserv.

06074332    \*\*Image available\*\*  
DEVICE AND METHOD FOR RETRIEVING SGML DOCUMENT

PUB. NO.:        11-015843 A]  
PUBLISHED:      January 22, 1999 (19990122)  
INVENTOR(s):    NAGAMURA EIJI  
APPLICANT(s):   TOSHIBA CORP  
APPL. NO.:      09-167284 [JP 97167284]  
FILED:          June 24, 1997 (19970624)  
INTL CLASS:     G06F-017/30

#### ABSTRACT

PROBLEM TO BE SOLVED: To provide an SGML document retrieving device capable of retrieving an SGML document recorded in a data base by an SQL to be a general relational data base language.

SOLUTION: At the time of inputting an SQL describing an SGML document retrieving instruction, an SQL analysis part 11 analyzes the SQL, separates the SQL into SQL information and SGML information and transfers the SGML information and the SQL analytical information respectively to an SGML analysis part 12 and a data access part 13. The SGML analysis part 12 extracts a tag identifier and a regrieving phrase from the SGML information and transfers the extracted contents to the data access part 13 as SGML analytical information. The data access part 13 detects an SGML document including the retrieving phrase in a tag sentence specified by the tag identifier and coincident with a condition specified by the SQL information from a user data base 17 and returns the detected result.

COPYRIGHT: (C) 1999, JPO  
?

(19)日本国特許庁(JP)

(12) 公開特許公報(A)

(11)特許出願公開番号

特開平11-15843

(43)公開日 平成11年(1999)1月22日

(51)Int.Cl.<sup>6</sup>

G 0 6 F 17/30

識別記号

F I

G 0 6 F 15/401

15/40

3 3 0 Z

3 7 0 A

審査請求 未請求 請求項の数7 O L (全 6 頁)

(21)出願番号

特願平9-167284

(22)出願日

平成9年(1997)6月24日

(71)出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72)発明者 永村 栄治

東京都青梅市末広町2丁目9番地 株式会

社東芝青梅工場内

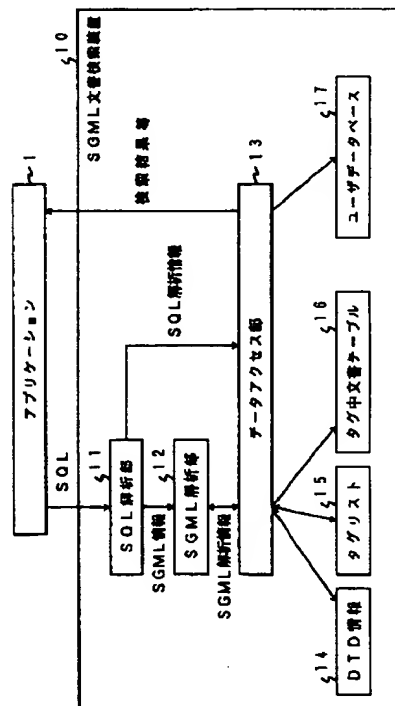
(74)代理人 弁理士 鈴江 武彦 (外6名)

(54)【発明の名称】 SGML文書検索装置およびSGML文書検索方法

(57)【要約】

【課題】データベースに登録されたSGML文書を汎用的な関係データベース言語であるSQLで検索可能とするSGML文書検索装置を提供する。

【解決手段】SQL解析部11は、SGML文書の検索指示を記述したSQLを入力すると、このSQLを解析してSQL情報とSGML情報とに分離し、SGML情報をSGML解析部12に、SQL解析情報をデータアクセス部13にそれぞれ引き渡す。SGML解析部12は、SGML情報からタグ識別子と検索語句とを抽出してSGML解析情報としてデータアクセス部13に引き渡す。そして、データアクセス部13は、SQL解析部11から受け取ったSQL解析情報とSGML解析部12から受け取ったSGML解析情報とから、タグ識別子で示されるタグの文中に検索語句を含み、かつSQL情報で指定される条件に合致するSGML文書をユーザデータベース17から検出し、その結果を返却する。



## 【特許請求の範囲】

【請求項 1】 データベースに登録された SGML 文書を検索する SGML 文書検索装置において、タグ識別子および検索語句を含む前記 SGML 文書の検索指示を記述した SQL を入力する SQL 入力手段と、前記 SQL 入力手段により入力された SQL を解析して SQL 情報と SGML 情報とに分離する SQL 解析手段と、

前記 SQL 解析手段により分離された SGML 情報からタグ識別子と検索語句とを得る SGML 解析手段と、前記 SGML 解析手段により得られたタグ識別子で示されるタグの文中に前記検索語句を含み、かつ前記 SQL 情報で指定される条件に合致する SGML 文書を前記データベースから検出するデータベース検索手段とを具備してなることを特徴とする SGML 文書検索装置。

【請求項 2】 前記 SQL 入力手段は、前記 SGML 文書の登録指示を記述した SQL を入力する手段を有し、前記 SGML 解析手段は、予め指定された DTD 情報に基づき、前記 SQL 解析手段により分離された SGML 情報からタグ文を得る手段を有し、少なくとも SGML 文書の識別番号または識別子およびタグ文を保持する検索テーブルと、前記登録指示された SGML 文書の識別番号または識別子および前記 SGML 解析手段により得られたタグ文を前記検索テーブルに登録するデータベース登録手段とをさらに具備してなることを特徴とする請求項 1 記載の SGML 文書検索装置。

【請求項 3】 前記データベース検索手段は、前記 SGML 解析手段により得られたタグ識別子で示されるタグの文中に前記検索語句を含む SGML 文書を前記検索テーブルによって特定する請求項 2 記載の SGML 文書検索装置。

【請求項 4】 データベースに登録された SGML 文書を検索する SGML 文書検索方法において、タグ識別子および検索語句を含む前記 SGML 文書の検索指示を記述した SQL を入力するステップと、前記入力された SQL を解析して SQL 情報と SGML 情報とに分離するステップと、前記分離された SGML 情報からタグ識別子と検索語句とを得るステップと、前記得られたタグ識別子で示されるタグの文中に前記検索語句を含み、かつ前記 SQL 情報で指定される条件に合致する SGML 文書を前記データベースから検出するステップとからなることを特徴とする SGML 文書検索方法。

【請求項 5】 SGML 文書の識別番号または識別子およびタグ文を保持する検索テーブルを有し、データベースに登録された SGML 文書を検索する SGML 文書検索方法において、タグ識別子および検索語句を含む前記 SGML 文書の検

索指示を記述した SQL を入力するステップと、前記入力された SQL を解析して SQL 情報と SGML 情報とに分離するステップと、前記分離された SGML 情報からタグ識別子と検索語句とを得るステップと、前記得られたタグ識別子で示されるタグの文中に前記検索語句を含み、かつ前記 SQL 情報で指定される条件に合致する SGML 文書を前記検索テーブルによって前記データベースから検出するステップとからなることを特徴とする SGML 文書検索方法。

【請求項 6】 データベースに登録された SGML 文書を検索するプログラムであって、タグ識別子および検索語句を含む前記 SGML 文書の検索指示を記述した SQL を入力し、前記入力された SQL を解析して SQL 情報と SGML 情報とに分離し、前記分離された SGML 情報からタグ識別子と検索語句とを得て、前記得られたタグ識別子で示されるタグの文中に前記検索語句を含み、かつ前記 SQL 情報で指定される条件に合致する SGML 文書を前記データベースから検出するようにコンピュータを動作させるプログラムを記録したコンピュータ読取り可能な記録媒体。

【請求項 7】 SGML 文書の識別番号または識別子およびタグ文を保持する検索テーブルを有するコンピュータ上でデータベースに登録された SGML 文書を検索するプログラムであって、タグ識別子および検索語句を含む前記 SGML 文書の検索指示を記述した SQL を入力し、前記入力された SQL を解析して SQL 情報と SGML 情報とに分離し、前記分離された SGML 情報からタグ識別子と検索語句とを得て、前記得られたタグ識別子で示されるタグの文中に前記検索語句を含み、かつ前記 SQL 情報で指定される条件に合致する SGML 文書を前記検索テーブルによって前記データベースから検出するように前記コンピュータを動作させるプログラムを記録したコンピュータ読取り可能な記録媒体。

## 【発明の詳細な説明】

【0001】

【発明の属する技術分野】 この発明は、データベースに登録された SGML (Standard Generalized Markup Language) 文書を汎用的な関係データベース言語である SQL で検索可能とする SGML 文書検索装置および SGML 文書検索方法に関する。

【0002】

【従来の技術】 近年、オフィスなどへのコンピュータの普及は目覚ましく、このオフィス環境で取り扱われる文

10

20

30

40

50

書の多くが、コンピュータ上で動作する文書作成ソフトウェアなどのアプリケーションプログラムで作成される程である。また、通信技術の向上に伴って、文書作成ソフトウェアなどで作成された電子文書をネットワークを介してコンピュータ相互間で直接に送受信するといったことが頻繁に行なわれるようになってきている。

【0003】このように、コンピュータで文書を取り扱う機会が増えたことから、コンピュータで解析が可能であり、かつ人間にとっても可読性のよいSGML文書が開発されるに至った。このSGML文書は、文書が  
10 章、節および段落などといった要素の木構造により表現されることに着目して開発されたものであり、その木構造に含まれる要素(タグ)の名前とそれらの構造上の関係とをDTD(Document Type Definition: 文書型定義)で定義することによって、ハードウェアおよびソフトウェアのいずれにも依存しない汎用的な電子文書として取り扱うことを可能としている。

【0004】ところで、従来、このSGML文書をデータベースに登録しておき、たとえばアプリケーションプログラムなどから検索するような場合には、まず、そのアプリケーションプログラムは、指定されたタグと語句とを文中に含むSGML文書をデータベースから取り出し、さらに、その取り出したSGML文書を解析して、指定されたタグ中にその語句が含まれている場合に、該当するSGML文書と判定するといったことを行なっていた。

【0005】

【発明が解決しようとする課題】しかしながら、このような方法では、アプリケーションプログラム自身がSGML文書の構文や内容を意識する必要があり、また、SGML文書の解析や、その解析結果とデータベースから返却される検索結果との比較など、アプリケーションプログラムの処理が複雑になるといった問題があった。

【0006】この発明はこのような実情に鑑みてなされたものであり、データベースに登録されたSGML文書を汎用的な関係データベース言語であるSQLで検索可能とするSGML文書検索装置およびSGML文書検索方法を提供することを目的とする。

【0007】

【課題を解決するための手段】この発明は、前述した目的を達成するために、少なくともSGML文書の識別番号または識別子およびタグ文を保持する検索テーブルを設け、登録指示されたSGML文書の識別番号または識別子およびこのSGML文書に含まれるタグ文をこの検索テーブルに登録しておき、たとえばアプリケーションプログラムなどからSGML文書の検索指示を受け取ったときに、その検索指示を記述したSQLを解析してSQL情報とSGML情報とに分離し、この分離されたSGML情報からタグ識別子と検索語句とを得て、その得

られたタグ識別子で示されるタグの文中に検索語句を含み、かつSQL情報で指定される条件に合致するSGML文書を、(1)検索テーブルから該当するSGML文書の識別番号を抽出し、(2)この検索テーブルから抽出した識別番号を検索キーにデータベースから検出するようにしたものである。

【0008】この発明によれば、たとえばアプリケーションプログラム自身はSGML文書の構成やタグの位置などを意識する必要がなくなり、SQLという標準インタフェースでSGML文書をタグ指定で検索指示することが可能となる。

【0009】

【発明の実施の形態】以下、図面を参照してこの発明の実施形態を説明する。図1は、この実施形態に係るSGML文書検索装置の機能ブロック図である。図1に示したように、このSGML文書検索装置10は、SQL解析部11、SGML解析部12およびデータアクセス部13の各処理部と、DTD情報14、タグリスト15、タグ中文書テーブル16およびユーザデータベース17の各データとを備えてなる。なお、この各処理部は、SGML文書検索装置10が構築されるコンピュータのCPUによって実行制御されるプログラムとして構成されるものであり、また、各データは、そのコンピュータが備える磁気ディスクやシステムメモリなどの記憶装置上に保持されるものである。

【0010】SQL解析部11は、アプリケーションプログラム1から送信されるSQLを解析してSGML情報とSQL解析情報とに分離し、SGML情報をSGML解析部12に、SQL解析情報をデータアクセス部13にそれぞれ引き渡すものである。SGML解析部12は、SQL解析部11から受け取ったSGML情報を解析し、このSGML情報からタグ識別子と検索語句とを抽出してSGML解析情報としてデータアクセス部13に引き渡すものである。そして、データアクセス部13は、SQL解析部11から受け取ったSQL解析情報とSGML解析部12から受け取ったSGML解析情報とに基づき、DTD情報14、タグリスト15、タグ中文書テーブル16およびユーザデータベース17へ各種データを登録したり、DTD情報14、タグリスト15およびタグ中文書テーブル16を参照してユーザデータベース17から該当するSGML文書を検出したりするのである。

【0011】DTD情報14は、ユーザデータベース17に登録されるSGML文書で用いられるタグの構造上の関係を保持し、また、タグリスト15は、SGML文書で用いられるタグを定義する。そして、タグ中文書テーブル16は、タグリスト15に基づいて作成され、ユーザデータベース17に登録されるSGML文書の各タグ中の文書を保持するものである。そして、ユーザデータベース17は、ユーザごとにユーザテーブルを設定

し、この設定したユーザテーブルにSGML文書の識別番号、登録日付および本文を格納するものである。

【0012】次に、この実施形態のSGML文書検索装置10の動作原理を「定義」、「登録」および「検索」の順に分説する。なお、このSGML文書検索装置10は、ユーザデータベース17に設定される図2に示したユーザテーブル(UT1)に対してSGML文書の検索を実行するものとする。ここで、本文はSGML文書の

```
ALTER TABLE UT1 ADD (
    文書 SGML(1000) KEY=番号
    DTDFILE=UT1.DTD TAGLIST=UT1.LST
)
```

このSQLは、SQL解析部11からデータアクセス部13に引き渡され、一方、データアクセス部13では、タグリスト15を参照し、このタグリスト15に定義されたタグをフィールドにもつタグ中文書テーブル16を作成する。ここでは、タグリスト15に「TAG1」および「TAG2」の2つのタグが定義されているものとする。その結果、データアクセス部13は、図3に示したタグ中文書テーブル(TGT1)を作成する。

#### 【0015】(2)登録

次に、図4乃至図7を参照してSGML文書の登録動作を説明する。ここでは、アプリケーションプログラム1からユーザデータベース17のユーザテーブル(UT1)に対して以下のようなSGML文書の登録指示を記述したSQLが送信されたものとする(図4の(1))。

```
【0016】
INSERT INTO UT1 VALUES (
    10, '19961015', SGML文書A
)
```

ここで、SGML文書Aは、図5に示すような構成であるものとする。

【0017】このSQLは、SQL解析部11でUT1の本文がSGML列属性であると認識され、SGML文書がSGML解析部12に引き渡される(図4の(2))。一方、このSGML文書(SGML文書A)

```
SELECT FROM UT1 WHERE
    日付 > 19960401 AND
    CONTAINS(本文, TAG('TAG2', 'コンピュータ'))
```

このSQLは、SQL解析部11でUT1の本文がSGML列属性であると認識され、SGML情報がSGML解析部12に引き渡される(図4の(2))。一方、このこのSGML情報を受け取ったSGML解析部12では、その内容を解析してタグ識別子と検索語句とを抽出し、データアクセス部13に引き渡す(図4の(3))。そして、データアクセス部13では、まず、タグ中文書テーブル(TGT1)16からタグ識別子

```
SELECT FROM UT1 日付 > 19960401 AND
    番号 IN (SELECT KEY FROM TKY1)
```

列であり、番号は一意的なキーであるとする。

#### 【0013】(1)定義

ここでは、図2に示したユーザテーブル(UT1)の文書列にSGML属性を付加するために、アプリケーションプログラム1から以下のSQLが送信されたものとする。

#### 【0014】

を受け取ったSGML解析部12では、その内容を解析してタグ識別子とタグ中文書とを抽出し、データアクセス部13に引き渡す。そして、データアクセス部13は、この受け取ったタグ識別子とタグ中文書とをタグ中文書テーブル(TGT1)16に格納する(図4の(3))。図6には、このときのタグ中文書テーブル(TGT1)16の格納状態が示されている。

20 【0018】また、データアクセス部13は、SQL解析部11から受け取ったSQL解析情報(登録データ)をユーザデータベース17のユーザテーブル(UT1)に登録する(図4の(4))。図7には、このときのユーザテーブル(UT1)の格納状態が示されている。

#### 【0019】(3)検索

次に、図8乃至図11を参照してSGML文書の検索動作を説明する。なお、このとき、ユーザデータベース17のユーザテーブル(UT1)は図9に示す状態にあり、また、タグ中文書テーブル(TGT1)16は図10に示す状態にあるものとする。そして、ここでは、アプリケーションプログラム1から以下のようなSQL(ユーザデータベース17のユーザテーブル(UT1)に登録されたSGML文書の検索指示を記述したSQL)が送信されたものとする(図8の(1))。

#### 【0020】

(TAG2)に検索語句('コンピュータ')を含むレコード集合を検索して図11に示すようなテンポラリキーリスト(TKY1)18を作成し、このテンポラリキーリスト(TKY1)18とユーザデータベース17のユーザテーブル(UT1)とを掛け合わせた結果(下記SQLの結果)をアプリケーションプログラム1に返却する(図4の(4))。

#### 【0021】

すなわち、この実施形態の SGML 文書検索装置 10 を適用すれば、アプリケーションプログラム 1 は SGML 文書の構成やタグの位置などを意識する必要がなくなり、SQL という標準インタフェースで SGML 文書をタグ指定で検索指示することが可能となる。

#### 【0022】

【発明の効果】以上詳述したように、この発明によれば、データベースに登録された SGML 文書を汎用的な関係データベース言語である SQL で検索できるようになるために、この SGML 文書を検索するアプリケーションプログラムなどは SGML 文書の構成やタグの位置などを意識する必要がなくなり、その処理を単純化することが可能となる。

#### 【図面の簡単な説明】

【図 1】この発明の実施形態に係る SGML 文書検索装置の機能ブロック図。

【図 2】同実施形態のユーザデータベースに設定されるユーザテーブル (UT1) の格納形式を示す図。

【図 3】同実施形態のタグ中文書テーブル (TGT1) の格納形式を示す図。

【図 4】同実施形態の SGML 文書の登録動作を説明する図。

【図 5】同実施形態の SGML 文書 A の構成を示す図。

【図 6】同実施形態のタグ中文書テーブル (TGT1) の格納状態を示す図。

【図 7】同実施形態のユーザテーブル (UT1) の格納状態を示す図。

【図 8】同実施形態の SGML 文書の検索動作を説明する図。

【図 9】同実施形態のユーザテーブル (UT1) の格納状態を示す図。

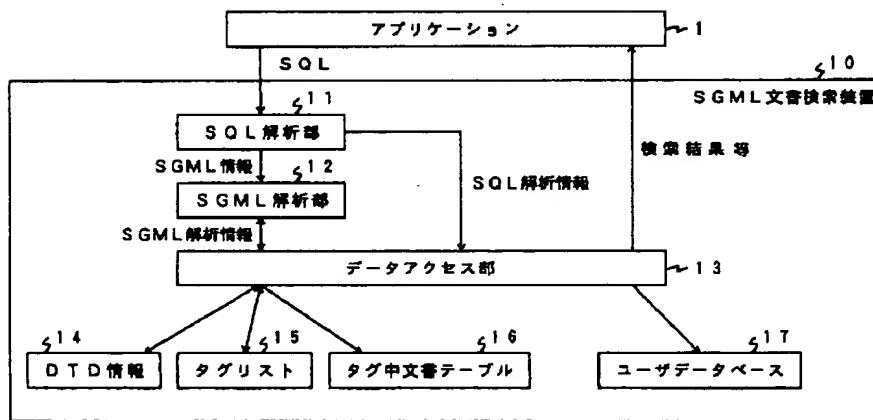
【図 10】同実施形態のタグ中文書テーブル (TGT1) の格納状態を示す図。

【図 11】同実施形態のテンポラリキーリスト (TKY1) の格納状態を示す図。

#### 【符号の説明】

10…SGML 文書検索装置、11…SQL 解析部、12…SGML 解析部、13…データアクセス部、14…DTD 情報、15…タグリスト、16…タグ中文書テーブル、17…ユーザデータベース。

【図 1】



【図 2】

ユーザテーブル (UT1)

番号	日付	本文

【図 6】

タグ中文書テーブル (TGT1)

KEY	TAG1	TAG2
10	山田一郎/田中二郎	最近のコンピュータ技術の進歩は、...

【図 5】

```
<TAG1>山田一郎</TAG1>
<TAG1>田中二郎</TAG1>
<TAG2>最近のコンピュータ技術の進歩
は、...</TAG2>
```

【図 11】

テンポラリキーリスト (TKY1)

KEY
10
15

【図 3】

タグ中文書テーブル (TGT1)

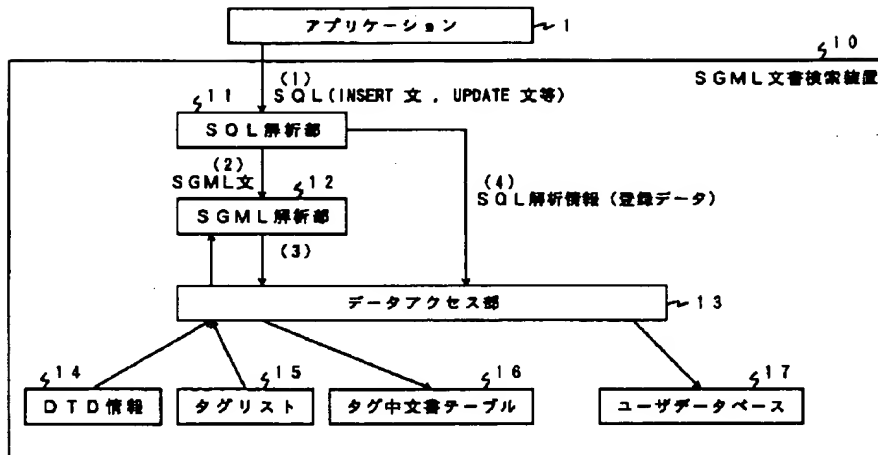
KEY	TAG1	TAG2

【図 7】

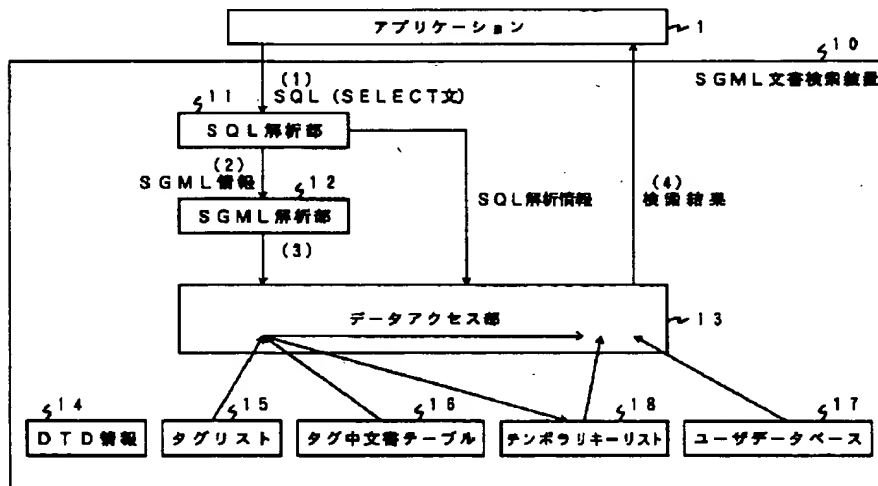
ユーザテーブル (UT1)

番号	日付	本文
10	19961015	<TAG1>山田一郎</TAG1> <TAG1>田中二郎</TAG1> <TAG2>最近のコンピュータ技術の進歩 は、...</TAG2>

【図 4】



【図 8】



【図 9】

ユーザテーブル (UT1)

番号	日付	本文
10	19981015	<TAG1>山田一郎</TAG1> <TAG1>田中二郎</TAG1> <TAG2>最近のコンピュータ技術の進歩は・・・</TAG2>
12	19980601	<TAG1>山田一郎</TAG1> <TAG1>田中二郎</TAG1> <TAG2>複数のプロセスやタスクを処理する・・・</TAG2>
15	19980105	<TAG1>中川三郎</TAG1> <TAG2>・・・コンピュータに対する要求が・・・</TAG2>

【図 10】

KEY	TAG1	TAG2
10	山田一郎/田中二郎	最近のコンピュータ技術の進歩は、・・・
12	山田一郎/田中二郎	複数のプロセスやタスクを処理する・・・
15	中川三郎	・・・コンピュータに対する要求が・・・